Міністерство освіти і науки України Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Фізико-технічний інститут

КРИПТОГРАФІЯ

Комп'ютерний практикум №1 «Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконав

студент групи ФБ-93

Флекевчук Данило

Перевірила

Селюх П.В.

Київ – 2021

Мета:

Засвоїти поняття ентропії та надлишковості мови. Навчитись проводити частотний аналіз тексту та за його результатами рахувати ентропію тексту. Покращити навички програмування.

Завдання:

- уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
- написати програму для підрахунку частот літер і частот біграм в тексті, а також підрахунку Н1, Н2 за безпосереднім означенням. Підрахувати частоти літер та біграм, а також значення Н1, Н2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1 Мб), де ймовірності заміняються відповідними частотами. Також отримати значення Н1, Н2 на тому ж самому тексті, в якому вилучено всі пробіли.
- за допомогою програми CoolPinkProgram оцінити значення H(10), H(20), H(30).
- використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи:

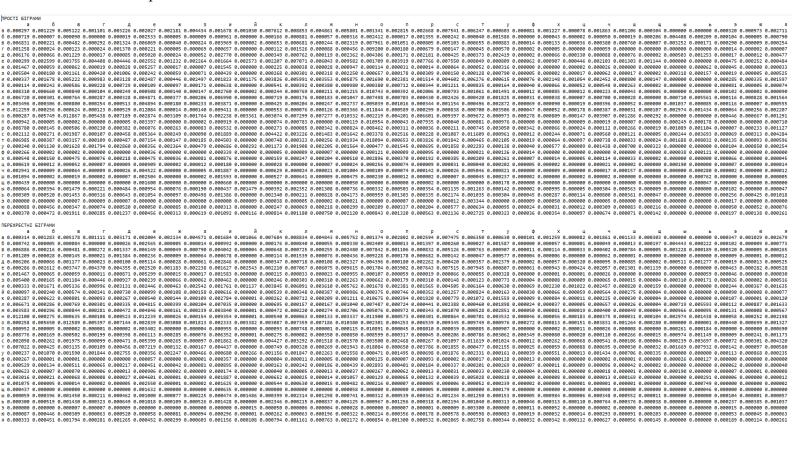
I.

В цій частині роботи я рахую частоту входження літер, біграм(двох типів) у текстовий файл, та за цими показниками рахуємо ентропію мови. Для обчислення було використано мову програмування Python, так як вона найкраще підходить для важких обчислень. Для підрахунку входжень я використовував модуль collections а саме Counter для підрахунку входжень у текст та OrderdDict для сортування. Я розбив текст на різні типи біграм, з допомогою циклів з різними кроками (перехрестні: 2, прості: 1). Код знаходится в тій же папці що і протокол.

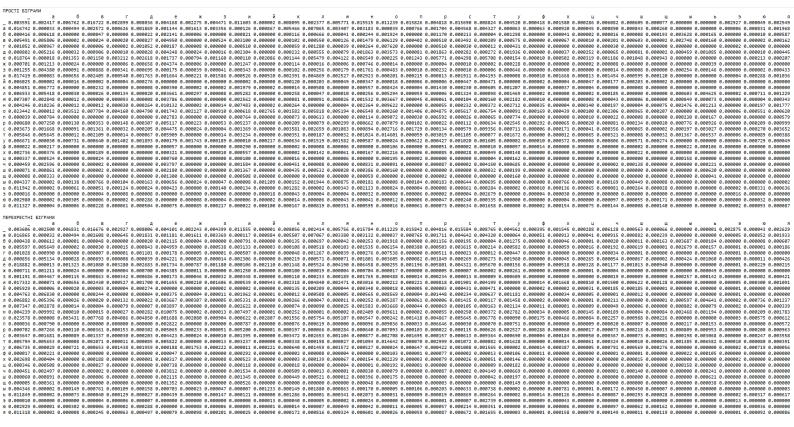
Далі я порахував ентропію для біграм та літер, і посортував літери за зростанням частоти. І вивів розраховані значення ентропії в консоль. Ентропію я рахував за формулою, що була наведена в завданні.

Для матриць з біграмами було виділено два файли, test_gap.txt та test.txt. Туди я повиводив значення частот для біграм у формі матриць, де на перехресті літер стоять біграми.

Без пробілів.



3 пробілами.



Літери.

_: 0.17018470956203863 a: 0.0660950216262635 a: 0.0660950216262635 6: 0.017392140247940356 6: 0.014432281039006059 B: 0.04626511057152555 B: 0.03839154172805269 r: 0.0168901357904256 r: 0.01401570956991704 Д: 0.032022662593665964 д: 0.02657292659675882 e: 0.08716245244012143 e: 0.07232882162468782 ж: 0.011410810541381544 ж: 0.009468876300924139 3: 0.01539836367908742 3: 0.012777812792978752 и: 0.06484995644547614 и: 0.0538135549307092944 й: 0.01001279339965014 и: 0.05381453410962033 к: 0.033031418595764174 к: 0.027410008745046444 л: 0.04596841762973196 л: 0.05400164658982581 0: 0.014173828122218503 0: 0.0952116931512893 л: 0.02749350028838554 10: 0.034710349880641936 0: 0.014828957705827764 0: 0.0439221927219154 7: 0.06474196021466327 7: 0.053723932277099797 y: 0.02966098677698897 y: 0.02461316956723838 ф: 0.00103449235837659117 ф: 0.00130578512396695 x: 0.00851034034240739 x: 0.007762018923965367 л: 0.002772298848119323 л: 0.00230049870399874	3 пробілами	Без пробілів
6:0.014432281039006059 B:0.03839154172805269 г:0.01401570956991704 д:0.032022662593665964 д:0.02657292659675882 e:0.08716245244012143 e:0.07232882162468782 ж:0.011410810541381544 ж:0.009468876300924139 3:0.01539836367908742 3:0.012777812792978752 и:0.06484995644547614 и:0.053813549307092944 й:0.01001279339965014 й:0.027410008745046444 л:0.03814534109620339 м:0.03814534109620339 м:0.0344351797128487 м:0.05400164658982581 o:0.11473828122218503 o:0.0952116931512893 п:0.027439350028838554 п:0.02276961923595081 р:0.041828957705827764 р:0.034710349880641936 c:0.05393020815976794 c:0.0439221927219154 т:0.0537233932277099797 y:0.02461316956723838 ф:0.0010330578512396695 x:0.007062018923965367 ц:0.00297298848119323 ц:0.002300498703999874 ч:0.018097082677641933 ч:0.018077253740279998 ш:0.002832635748888589 ш:0.00286646251053737 ы:0.013700572761149934 ы:0.01370057276149934 ы:0.013700572761149934 ы:0.013616990774036282 ы:0.0004661070362170978 я:0.021361891809138618	_: 0.17018470956203863	a: 0.07965137392567485
B: 0.03839154172805269 г: 0.01401570956991704 д: 0.02657292659675882 д: 0.032022662593665964 е: 0.07232882162468782 ж: 0.011410810541381544 ж: 0.009468876300924139 3: 0.01539836367908742 з: 0.012777812792978752 н: 0.06484995644547614 н: 0.053813549307092944 н: 0.01001279339965014 н: 0.03830877892365023 к: 0.033031418595764174 н: 0.03814534109620339 м: 0.03144351797128487 м: 0.05400164658982581 о: 0.11473828122218503 п: 0.024616931512893 п: 0.02749393500288388554 п: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 с: 0.05393020815976794 с: 0.0439221927219154 т: 0.06474196021466327 т: 0.053723932277099797 y: 0.02461316956723838 ф: 0.0010330578512396695 x: 0.0007062018923965367 п: 0.002772298848119323 п: 0.002300498703999874 п: 0.002772298848119323 п: 0.002300498703999874 п: 0.002300498703999874 п: 0.002300498703999874 п: 0.002300498703999874 п: 0.002772298848119323 п: 0.002300498703999874 п: 0.00248268646251053737 р: 0.003232037639653368 р: 0.0019268646251053737 р: 0.00352	a: 0.0660950216262635	б: 0.017392140247940356
г: 0.01401570956991704 д: 0.02657292659675882 е: 0.07232882162468782 ж: 0.009468876300924139 д: 0.0253813549307092944 д: 0.06484995644547614 й: 0.00830877892365023 к: 0.033031418595764174 й: 0.0027410008745046444 д: 0.04596841762973196 л: 0.053814534109620339 м: 0.033044351797128487 м: 0.02609234296338898 н: 0.05400164658982581 п: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 с: 0.0439221927219154 т: 0.031303578512396695 д: 0.0010330578512396695 х: 0.002300498703999874 д: 0.00296098677698897 д: 0.0024826687171568357 д: 0.00322037639653368 д: 0.013700572761149934 д: 0.00322037639653368 д: 0.013700572761149934 д: 0.00325288830899 д: 0.0029258483088971 д: 0.0029258483088971 д: 0.004661070362170978 д: 0.0021361891809138618	б : 0.014432281039006059	в: 0.04626511057152555
д: 0.02657292659675882 e: 0.07232882162468782 ж: 0.009468876300924139 ж: 0.011410810541381544 ж: 0.012777812792978752 и: 0.06484995644547614 и: 0.053813549307092944 й: 0.01001279339965014 и: 0.027410008745046444 и: 0.033031418595764174 и: 0.03814534109620339 м: 0.03144351797128487 м: 0.02409234296338898 н: 0.05400164658982581 и: 0.05400164658982581 о: 0.011473828122218503 и: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 c: 0.052930020815976794 c: 0.0439221927219154 т: 0.06474196021466327 y: 0.02461316956723838 ф: 0.0010330578512396695 x: 0.0070620189239965367 ц: 0.002300498703999874 ч: 0.015017253740279998 ш: 0.008323635748888589 ш: 0.002300498703999874 ц: 0.002300498703999874 ч: 0.0150752761149934 р: 0.016510368824929803 р: 0.013700572761149934 р: 0.002322037639653368 р: 0.0029258483088971 р: 0.0029258483088971 р: 0.004661070362170978 р: 0.021361891809138618	в: 0.03839154172805269	г: 0.0168901357904256
e: 0.07232882162468782 ж: 0.001410810541381544 ж: 0.009468876300924139 3: 0.01539836367908742 3: 0.012777812792978752 и: 0.06484995644547614 и: 0.053813549307092944 й: 0.01001279339965014 й: 0.00830877892365023 к: 0.033031418595764174 к: 0.027410008745046444 л: 0.04596841762973196 л: 0.03814534109620339 м: 0.03144351797128487 м: 0.05400164658982581 о: 0.05400164658982581 о: 0.0952116931512893 л: 0.02276961923595081 р: 0.034710349880641936 с: 0.052930020815976794 с: 0.0439221927219154 т: 0.06474196021466327 т: 0.053723932277099797 y: 0.02966098677698897 у: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 x: 0.00851034034240739 x: 0.007062018923965367 лг: 0.002300498703999874 лг: 0.002300498703999874 лг: 0.002306482687171568357 лг: 0.002482687171568357 лг: 0.013700572761149934 лг: 0.002320237639653368 лг: 0.019268646251053737 лг: 0.00352588989202750462 лг: 0.004661070362170978 лг: 0.0021361891809138618	г: 0.01401570956991704	д: 0.032022662593665964
ж: 0.009468876300924139 3: 0.012777812792978752 и: 0.06484995644547614 и: 0.053813549307092944 й: 0.01001279339965014 й: 0.00830877892365023 к: 0.033031418595764174 к: 0.027410008745046444 л: 0.04596841762973196 л: 0.03814534109620339 м: 0.03144351797128487 м: 0.02609234296338898 н: 0.05400164658982581 o: 0.11473828122218503 o: 0.0952116931512893 п: 0.027439350028838554 п: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 c: 0.052930020815976794 c: 0.0439221927219154 т: 0.06474196021466327 т: 0.053723932277099797 y: 0.02966098677698897 y: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 x: 0.00851034034240739 x: 0.007062018923965367 ц: 0.002300498703999874 ч: 0.018097082677641933 ц: 0.002482687171568357 ы: 0.015017253740279998 ш: 0.0029918516250465807 ш: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.002322037639653368 ь: 0.019268646251053737 э: 0.002361891809138618	д: 0.02657292659675882	e: 0.08716245244012143
3: 0.012777812792978752 и: 0.053813549307092944 й: 0.00830877892365023 к: 0.033031418595764174 к: 0.027410008745046444 л: 0.04596841762973196 л: 0.03814534109620339 м: 0.03144351797128487 м: 0.02609234296338898 н: 0.06507662985300645 н: 0.05400164658982581 о: 0.11473828122218503 о: 0.0952116931512893 п: 0.027439350028838554 п: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 с: 0.052930020815976794 с: 0.0439221927219154 т: 0.06474196021466327 т: 0.053723932277099797 y: 0.02966098677698897 у: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 x: 0.00851034034240739 x: 0.007062018923965367 ц: 0.002772298848119323 ц: 0.002300498703999874 ч: 0.018097082677641933 ч: 0.015017253740279998 ш: 0.008232635748888589 ш: 0.00683157513255442 ш: 0.002918516250465807 ш: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.0035258989202750462 э: 0.0029258483088971 ю: 0.006616990774036282	e: 0.07232882162468782	ж: 0.011410810541381544
и : 0.053813549307092944 й : 0.01001279339965014 й : 0.00830877892365023 к : 0.033031418595764174 к : 0.027410008745046444 л : 0.04596841762973196 л : 0.03814534109620339 м : 0.03144351797128487 м : 0.05400164658982581 о : 0.11473828122218503 о : 0.0952116931512893 л : 0.027439350028838554 п : 0.02276961923595081 р : 0.041828957705827764 р : 0.034710349880641936 с : 0.052930020815976794 с : 0.0439221927219154 т : 0.06474196021466327 т : 0.053723932277099797 у : 0.02966098677698897 у : 0.02461316956723838 ф : 0.0012449235837659117 ф : 0.0010330578512396695 х : 0.00851034034240739 х : 0.007062018923965367 ц : 0.002772298848119323 ц : 0.002300498703999874 ч : 0.018097082677641933 ч : 0.015017253740279998 ш : 0.008232635748888589 ш : 0.002482687171568357 ы : 0.016510368824929803 ы : 0.013700572761149934 ь : 0.02322037639653368 ь : 0.019268646251053737 э : 0.0035258989202750462 э : 0.0029258483088971 ю : 0.005616990774036282 ю : 0.004661070362170978 я : 0.021361891809138618	ж: 0.009468876300924139	з: 0.01539836367908742
й: 0.00830877892365023 к: 0.033031418595764174 к: 0.027410008745046444 л: 0.04596841762973196 л: 0.03814534109620339 м: 0.03144351797128487 м: 0.02609234296338898 н: 0.06507662985300645 н: 0.05400164658982581 о: 0.11473828122218503 о: 0.0952116931512893 п: 0.027439350028838554 р: 0.034710349880641936 с: 0.052930020815976794 с: 0.0439221927219154 т: 0.06474196021466327 т: 0.053723932277099797 у: 0.02966098677698897 у: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 х: 0.00851034034240739 х: 0.007062018923965367 ц: 0.002772298848119323 ц: 0.002300498703999874 ч: 0.018097082677641933 ч: 0.015017253740279998 ш: 0.008232635748888589 ш: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.0035258989202750462 э: 0.0029258483088971 ю: 0.005616990774036282 ю: 0.004661070362170978 я: 0.021361891809138618	з: 0.012777812792978752	и: 0.06484995644547614
к: 0.027410008745046444 л: 0.04596841762973196 л: 0.03814534109620339 м: 0.03144351797128487 м: 0.02609234296338898 н: 0.06507662985300645 н: 0.05400164658982581 о: 0.11473828122218503 о: 0.0952116931512893 п: 0.027439350028838554 п: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 с: 0.052930020815976794 с: 0.0439221927219154 т: 0.06474196021466327 т: 0.053723932277099797 у: 0.02966098677698897 у: 0.02461316956723838 ф: 0.0012349235837659117 ф: 0.0010330578512396695 х: 0.00851034034240739 х: 0.007062018923965367 ц: 0.002772298848119323 ц: 0.002300498703999874 ч: 0.018097082677641933 ч: 0.015017253740279998 ш: 0.008232635748888589 ш: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.0035258989202750462 э: 0.0029258483088971 ю: 0.005616990774036282 ю: 0.004661070362170978 я: 0.021361891809138618	и: 0.053813549307092944	й: 0.01001279339965014
л: 0.03814534109620339 м: 0.03144351797128487 м: 0.02609234296338898 н: 0.05400164658982581 о: 0.11473828122218503 о: 0.0952116931512893 п: 0.027439350028838554 п: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 с: 0.052930020815976794 т: 0.06474196021466327 р: 0.0439221927219154 т: 0.06474196021466327 у: 0.02966098677698897 р: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 х: 0.00851034034240739 х: 0.007062018923965367 п: 0.002300498703999874 ч: 0.018097082677641933 ч: 0.015017253740279998 пг: 0.002482687171568357 пг: 0.002482687171568357 пг: 0.0013700572761149934 р: 0.013700572761149934 р: 0.02322037639653368 р: 0.019268646251053737 р: 0.0035258989202750462 р: 0.0029258483088971 р: 0.004661070362170978 п: 0.021361891809138618	й: 0.00830877892365023	к: 0.033031418595764174
M: 0.02609234296338898 H: 0.05400164658982581 o: 0.11473828122218503 o: 0.0952116931512893 п: 0.027439350028838554 п: 0.02276961923595081 p: 0.041828957705827764 p: 0.034710349880641936 c: 0.052930020815976794 c: 0.0439221927219154 T: 0.06474196021466327 T: 0.053723932277099797 y: 0.02966098677698897 y: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 x: 0.00851034034240739 x: 0.007062018923965367 ц: 0.002772298848119323 ц: 0.002300498703999874 ч: 0.018097082677641933 ч: 0.015017253740279998 ш: 0.008232635748888589 ш: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.0025888899202750462 э: 0.0029258483088971 ю: 0.005616990774036282 ю: 0.004661070362170978 я: 0.021361891809138618	к: 0.027410008745046444	л: 0.04596841762973196
H: 0.05400164658982581 0: 0.11473828122218503 0: 0.0952116931512893 п: 0.027439350028838554 п: 0.02276961923595081 p: 0.041828957705827764 p: 0.034710349880641936 c: 0.052930020815976794 c: 0.0439221927219154 T: 0.06474196021466327 T: 0.053723932277099797 y: 0.02966098677698897 y: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 x: 0.00851034034240739 x: 0.002300498703999874 प: 0.018097082677641933 प: 0.002300498703999874 प: 0.018097082677641933 पा: 0.002482687171568357 पा: 0.0029918516250465807 पा: 0.002482687171568357 पा: 0.013700572761149934 b: 0.019268646251053737 9: 0.0035258989202750462 9: 0.0029258483088971 0: 0.005616990774036282 10: 0.004661070362170978 9: 0.021361891809138618	л: 0.03814534109620339	м: 0.03144351797128487
о: 0.0952116931512893 п: 0.027439350028838554 п: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 с: 0.052930020815976794 с: 0.0439221927219154 т: 0.06474196021466327 т: 0.053723932277099797 у: 0.02966098677698897 у: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 х: 0.00851034034240739 х: 0.002300498703999874 ц: 0.0022772298848119323 ч: 0.015017253740279998 ш: 0.008232635748888589 ш: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.0035258989202750462 э: 0.0029258483088971 ю: 0.004661070362170978	м: 0.02609234296338898	н: 0.06507662985300645
п: 0.02276961923595081 р: 0.041828957705827764 р: 0.034710349880641936 с: 0.052930020815976794 с: 0.0439221927219154 т: 0.06474196021466327 т: 0.053723932277099797 у: 0.02966098677698897 у: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 х: 0.00851034034240739 х: 0.007062018923965367 ц: 0.002772298848119323 ц: 0.002300498703999874 ч: 0.018097082677641933 ч: 0.015017253740279998 ш: 0.008232635748888589 ш: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.0035258989202750462 э: 0.0029258483088971 ю: 0.005616990774036282 ю: 0.004661070362170978 я: 0.021361891809138618	н: 0.05400164658982581	o: 0.11473828122218503
р: 0.034710349880641936 c: 0.0439221927219154 т: 0.053723932277099797 у: 0.02461316956723838 ф: 0.0010330578512396695 х: 0.007062018923965367 ц: 0.002300498703999874 ч: 0.015017253740279998 ш: 0.00683157513255442 щ: 0.002482687171568357 ы: 0.013700572761149934 ь: 0.019268646251053737 э: 0.0029258483088971 ю: 0.004661070362170978 c: 0.052930020815976794 т: 0.06474196021466327 у: 0.02966098677698897 у: 0.02966098677698897 у: 0.02966098677698897 у: 0.0012449235837659117 х: 0.00851034034240739 п: 0.002772298848119323 п: 0.018097082677641933 п: 0.008232635748888589 п: 0.0029918516250465807 п: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ы: 0.0035258989202750462 п: 0.0035258989202750462 п: 0.005616990774036282 п: 0.004661070362170978	o: 0.0952116931512893	п: 0.027439350028838554
c: 0.0439221927219154 T: 0.06474196021466327 т: 0.053723932277099797 y: 0.02966098677698897 y: 0.02461316956723838 ф: 0.0012449235837659117 ф: 0.0010330578512396695 x: 0.00851034034240739 x: 0.002300498703999874 ц: 0.018097082677641933 ч: 0.015017253740279998 ш: 0.008232635748888589 ш: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.0035258989202750462 э: 0.0029258483088971 ю: 0.005616990774036282 ю: 0.004661070362170978 я: 0.021361891809138618	п: 0.02276961923595081	p: 0.041828957705827764
$\begin{array}{llllllllllllllllllllllllllllllllllll$	p: 0.034710349880641936	c: 0.052930020815976794
$\begin{array}{llllllllllllllllllllllllllllllllllll$	c: 0.0439221927219154	т: 0.06474196021466327
$\begin{array}{llllllllllllllllllllllllllllllllllll$	т: 0.053723932277099797	y: 0.02966098677698897
$\begin{array}{llllllllllllllllllllllllllllllllllll$	y: 0.02461316956723838	ф: 0.0012449235837659117
$\begin{array}{llllllllllllllllllllllllllllllllllll$	ф: 0.0010330578512396695	x: 0.00851034034240739
$\begin{array}{lll} \mathbf{u}: 0.015017253740279998 & \mathbf{m}: 0.008232635748888589 \\ \mathbf{m}: 0.00683157513255442 & \mathbf{m}: 0.0029918516250465807 \\ \mathbf{m}: 0.002482687171568357 & \mathbf{m}: 0.016510368824929803 \\ \mathbf{b}: 0.013700572761149934 & \mathbf{b}: 0.02322037639653368 \\ \mathbf{b}: 0.019268646251053737 & \mathbf{g}: 0.0035258989202750462 \\ \mathbf{g}: 0.0029258483088971 & \mathbf{g}: 0.005616990774036282 \\ \mathbf{g}: 0.004661070362170978 & \mathbf{g}: 0.021361891809138618 \\ \end{array}$	x: 0.007062018923965367	ц: 0.002772298848119323
$\begin{array}{llllllllllllllllllllllllllllllllllll$	ц: 0.002300498703999874	ч: 0.018097082677641933
щ: 0.002482687171568357 ы: 0.016510368824929803 ы: 0.013700572761149934 ь: 0.02322037639653368 ь: 0.019268646251053737 э: 0.0035258989202750462 э: 0.0029258483088971 ю: 0.005616990774036282 ю: 0.004661070362170978 я: 0.021361891809138618	ч: 0.015017253740279998	ш: 0.008232635748888589
ы : 0.013700572761149934 ь : 0.019268646251053737 э : 0.0029258483088971 ю : 0.004661070362170978 ь : 0.02322037639653368 э : 0.0035258989202750462 ю : 0.005616990774036282 я : 0.021361891809138618	ш: 0.00683157513255442	щ: 0.0029918516250465807
ь : 0.019268646251053737 э : 0.0035258989202750462 э : 0.0029258483088971 ю : 0.005616990774036282 ю : 0.004661070362170978 я : 0.021361891809138618	· ·	ы: 0.016510368824929803
э : 0.0029258483088971 ю : 0.005616990774036282 ю : 0.004661070362170978 я : 0.021361891809138618	ы: 0.013700572761149934	ь: 0.02322037639653368
ю: 0.004661070362170978 я: 0.021361891809138618	ь: 0.019268646251053737	э: 0.0035258989202750462
	э: 0.0029258483088971	ю: 0.005616990774036282
	ю: 0.004661070362170978	я: 0.021361891809138618
я: 0.017726445493149712	я: 0.017726445493149712	

Біграми (перші 10 в кожній категорії).

3 пробілами		Без пробілів	
Перехресні	Прості	Перехресні	Прості
o_: 0.011721	o_: 0.011721	то: 0.018084	то: 0.009047
e_: 0.009382	e_: 0.009382	ов: 0.012589	ов: 0.006433
и_: 0.008710	и_: 0.008710	не: 0.012239	на: 0.006130
a_: 0.008371	a_: 0.008371	на: 0.012100	не: 0.006042
_в: 0.008361	_в: 0.008361	но: 0.011900	но: 0.005922
_п: 0.007913	_п: 0.007913	ст: 0.011619	ст: 0.005805
_c: 0.007845	_c: 0.007845	по: 0.010891	по: 0.005477
_н: 0.007757	_н: 0.007757	ко: 0.010675	ко: 0.005371
то: 0.007300	то: 0.007300	он: 0.010380	он: 0.005166
ь_: 0.005971	ь_: 0.005971	от: 0.009749	от: 0.004836

Без пробілв

Ентропія	Значення ентропії Н	Надлишковість R
Літери	4.3497	13.00%
Біграми	3.9496	21.06%
Перехресні біграми	3.9502	21.00%

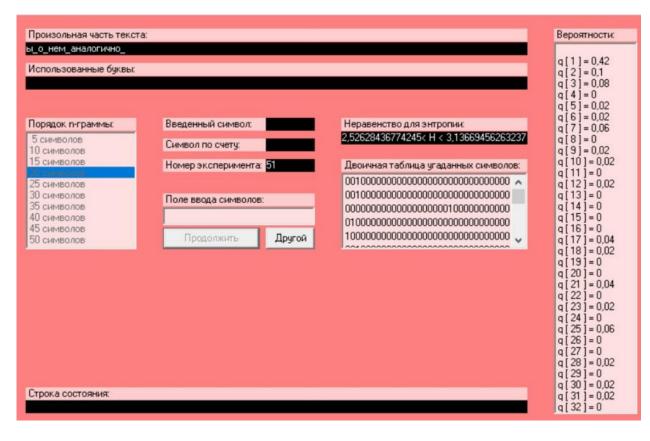
3 пробілами

Ентропія	Значення ентропії Н	Надлишковість R
Літери	2.1749	56.41%
Біграми	3.9496	21.06%
Перехресні біграми	3.9502	21.00%

Тут видно, що значення ентропії дуже сильно падає на літерах порівняно з іншими експериментами, бо пробіли дуже часто трапляюся у тексті.

II.

Для наступного завдання я використовую надану мені CoolPinkProgram. З допомогою цієї програми нам потрібно порахувати ентропію російської мови, а з нею і її надлишковість H(10), H(20), H(30).



Произольная часть текста	a.		Вероятности:
временами	-		Боролиновия.
Использованные буквы:			q[1]=0,4 q[2]=0,06 q[3]=0,08 q[4]=0,08 q[5]=0,04
Порядок п-граммы:	Введенный символ:	Неравенство для энтропии:	q[6]=0 q[7]=0
5 символов	Символ по счету:	2,74437798355353< H < 3,25998720177651	q[8]=0 q[9]=0
15 символов 20 символов 25 символов	Номер эксперимента: 51	Двоичная таблица угаданных символов: 000000000000000000000000000000000000	q[10] = 0 q[11] = 0 q[12] = 0
30 символов 35 символов 40 символов	Поле ввода символов:	10000000000000000000000000000000000000	q[13]=0 q[14]=0 q[15]=0,04
45 символов 50 символов	Продолжить Другой	010000000000000000000000000000000000000	q[16] = 0,04 q[17] = 0 q[18] = 0,02
			q[19]=0 q[20]=0,02 q[21]=0
			q[22]=0,04 q[23]=0,02
			q[24]=0 q[25]=0,04
			q[26]=0,04 q[27]=0,02 q[28]=0
Строка состояния:			q[29]=0 q[30]=0,02 q[31]=0,02
			q[32]=0,02

Произольная часть текста:			Вероятности:
ещи_как_бы_вам_понрави. Использованные буквы: Порядок n-граммы:	лось_ес Введенный символ:	Неравенство для энтропии:	q[1]=0.5686274 q[2]=0.1568627 q[3]=0.0784313 q[4]=0 q[5]=0 q[6]=0 q[7]=0
5 символов 10 символов 15 символов 20 символов 25 символов	Символ по счету: Номер эксперимента: 52	1,64952187862109 H < 2,24337732513425 Двоичная таблица угаданных символов: 1000000000000000000000000000000000000	q[7]=0 q[8]=0 q[9]=0 q[10]=0 q[11]=0 q[12]=0
35 символов 40 символов 45 символов 50 символов	Продолжить Другой	10000000000000000000000000000000000000	q[13] = 0,019607 q[14] = 0,019607 q[15] = 0 q[16] = 0,019607 q[17] = 0 q[18] = 0,019607 q[19] = 0,039215 q[20] = 0 q[21] = 0 q[22] = 0 q[23] = 0 q[23] = 0
Строка состояния:			q[25] = 0,019607 q[26] = 0 q[27] = 0,019607 q[28] = 0 q[29] = 0 q[30] = 0,019607 q[31] = 0 q[32] = 0

H(10) = 3; H(20) = 2,8314; H(30) = 1,9464

Для обрахунку ентропії, я просумував усі значення ентропій і поділив їх на 3.

$$H = (3 + 2,8314 + 1.9464) \backslash 3 = 2,5926$$

Ентропія	Значення ентропії	Надлишковість
H(10)	3.0000	40.00%
H(20)	2.8140	43.72%
H(30)	1.9460	61.08%
Н	2,5926	48.00%

Надлишковість рахується за формулою $1 - (H\backslash H(0))$, де H(0) для російської мови рівне 5. У мене це вийшло 0,48, я думаю такт мала надлишковість зумовлена моїми неякісними знаннями російської мови.

Висновок:

В ході роботи я навчився обраховувати ентропію джерела тексту для символів та біграм, також я навився обраховувати надлишковість тексту. Для російської мови надлишковість склала 48%, за вказаними вище причинами. Для пошуку ентропії біграм та літер було обрано "Злочин і кара". Якщо не забрати пробіли то, тоді вони будуть найбільш вживаним символом ("_": 172811), якщо ні то це літери: "о": 96681, "е": 73380 та "а": 67116 відповідно, а най менш вживані: "ц": 2336, "ф": 1049. Найчастішими біграмами з пробілом виявились: "о_": 0.023578, "е_": 0.018882, "и_": 0.017332, а без пробілів: "то": 0.018084, "ов": 0.012589, "не": 0.012239. Наявність пробілів дуже полегшує роботу криптоаналітика.